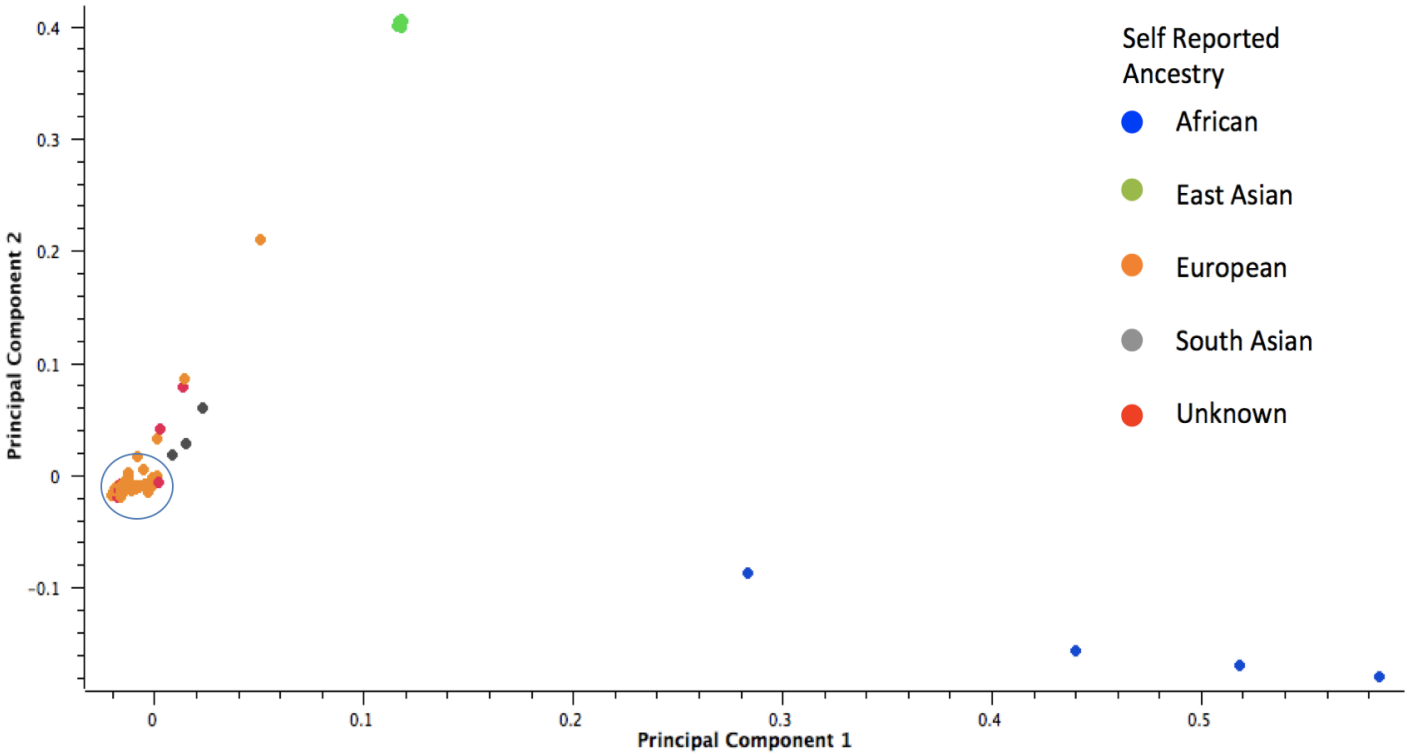
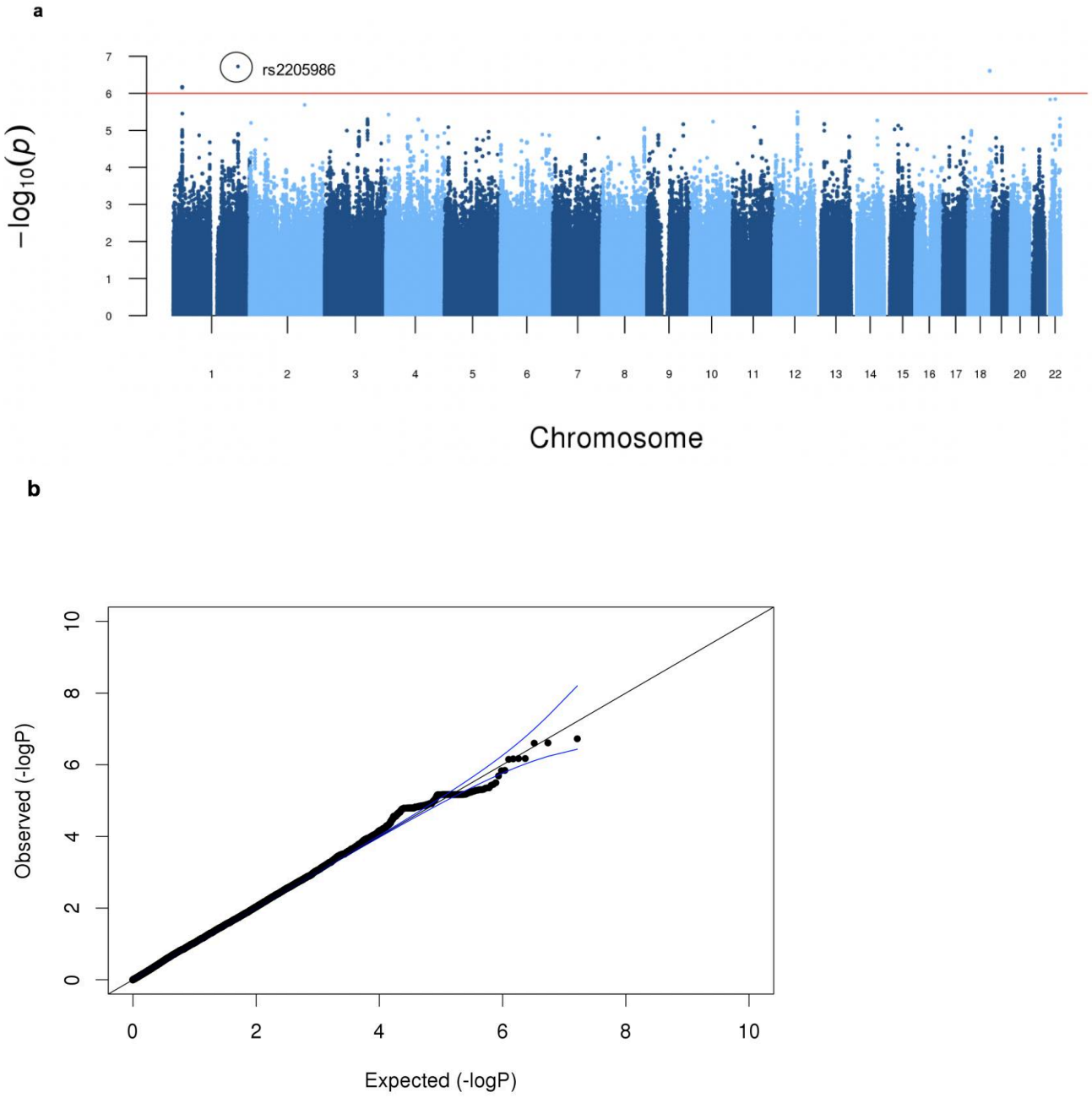


Supplementary Figure 1

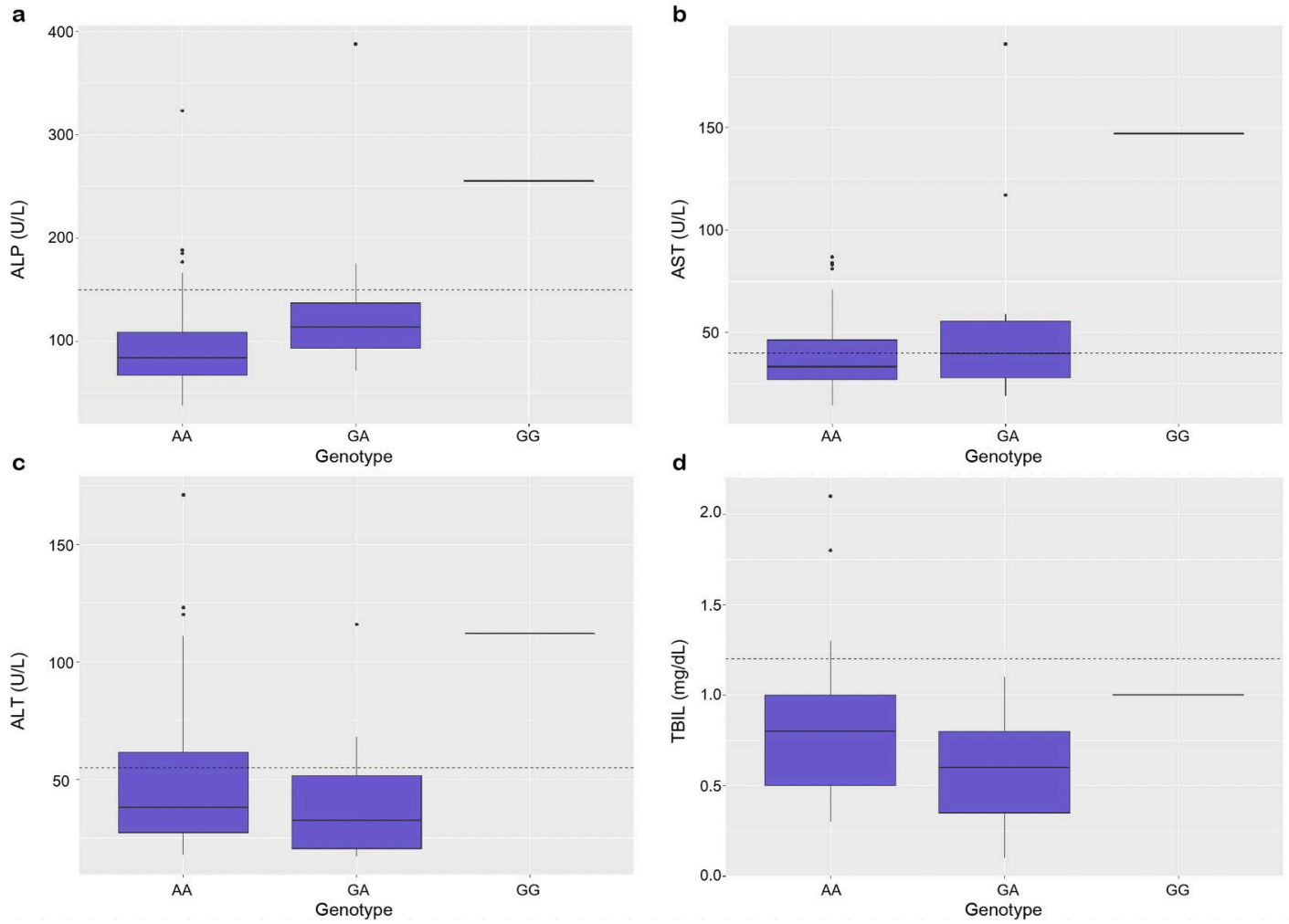
Genetic ancestry determined by principal-components analysis of patients genotyped on the Illumina MEGA array.

The first two principal components are plotted to visualize the distribution of genetic ancestry. Each point represents one patient, colored by their self-reported ancestry (*n* = 170 in stage 1 and *n* = 10 in stage 2). The circled patients correspond to those that were considered to be of European genetic ancestry for analyses.



Supplementary Figure 2

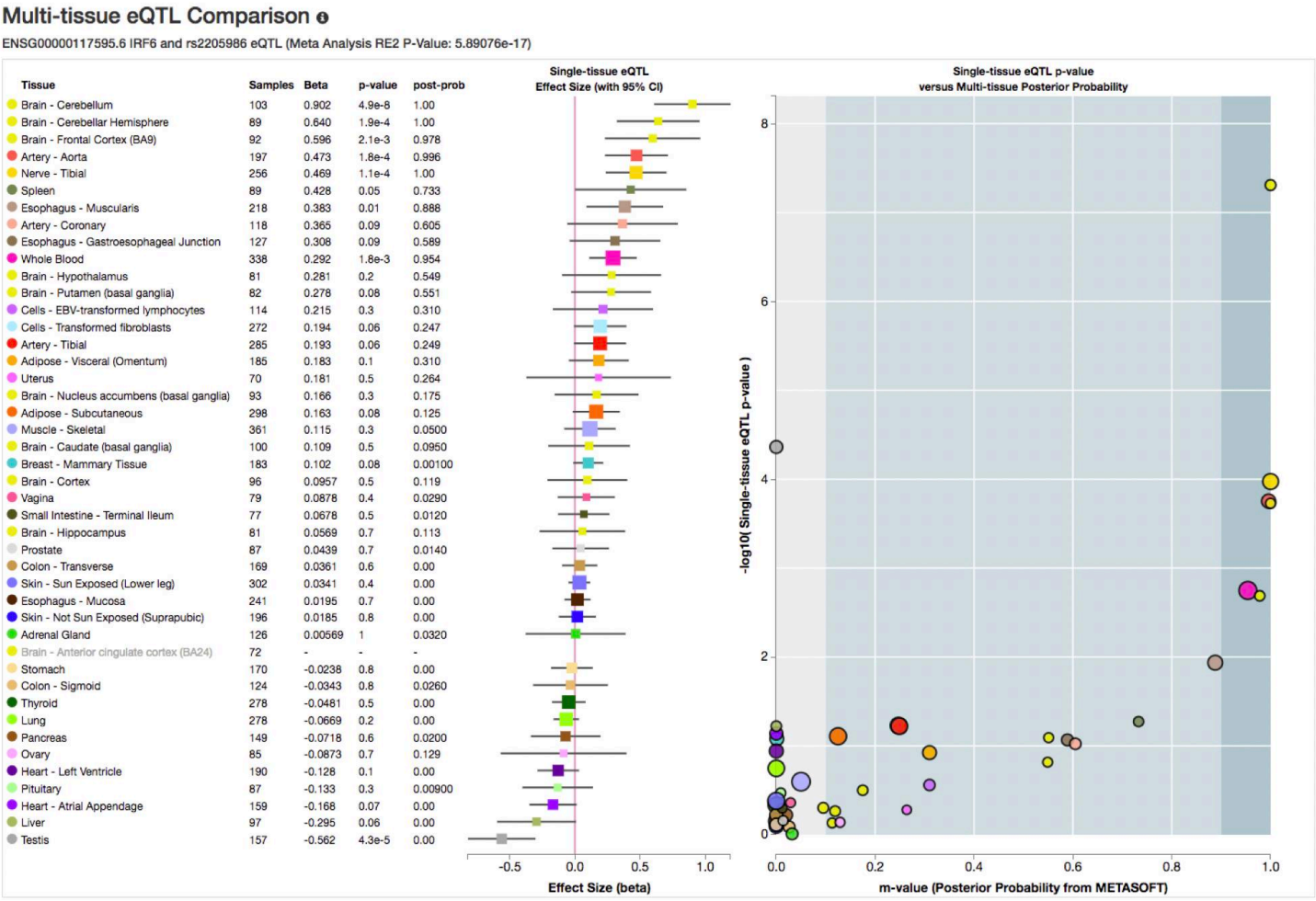
Genome-wide scan of interferon-β-induced liver injury in multiple sclerosis.
a, Manhattan plot showing the observed distribution of $-\log_{10}$ (*P* values) against the chromosomal location (GRCh37) of genetic variants in the imputed stage 1 cohort of European genetic ancestry (*n* = 151). *P* values are from logistic regression employing an additive model with adjustment for MS disease course at the start of IFN-β, with the variant most strongly associated with liver injury from IFN-β (rs2205986) circled. The red horizontal line ($P = 1.0 \times 10^{-6}$) indicates the screening threshold employed to prioritize variants for subsequent stage 2 analyses. **b**, Quantile–quantile plot showing the distributions of adjusted observed $-\log_{10}$ (*P* values) plotted against expected $-\log_{10}$ (*P* values) in the stage 1 cohort of patients with European genetic ancestry (*n* = 151). A genomic inflation factor (λ_{GC}) of 1.06 indicates no obvious population stratification. The blue lines indicate the 95% confidence interval, and *P* values were obtained in the same manner as stated in **a**.



Supplementary Figure 3

Box plots of peak biochemical liver test results from the BioVU multiple sclerosis cohort, stratified by rs2205986 genotype status.

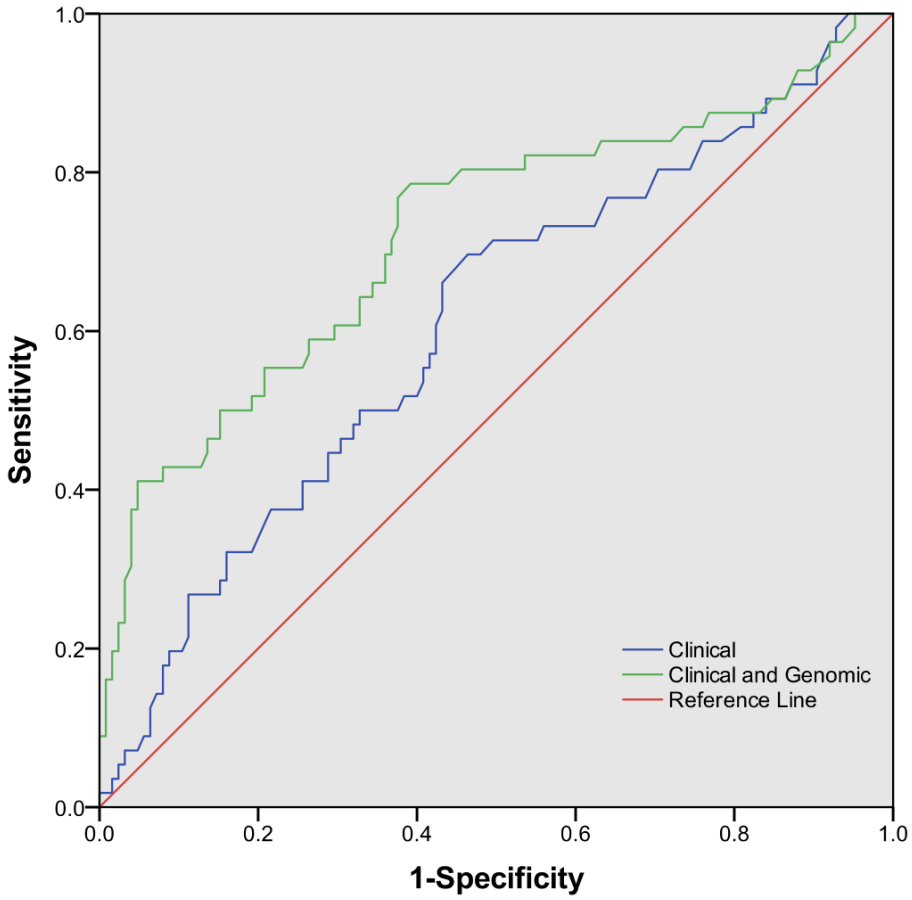
a–**d**, Peak alkaline phosphatase (ALP) (**a**) and aspartate aminotransferase (AST) (**b**) values were statistically significant ($P = 4.9 \times 10^{-4}$ for ALP and $P = 7.6 \times 10^{-5}$ for ASP) with carriers displaying elevated levels, while no association was detected for mean peak alanine aminotransferase (ALT) (**c**) and total bilirubin (TBIL) (**d**) values. Each box plot summarizes the relevant biochemical liver test result by rs2205986 genotype for MS patients ($n = 87$) exposed to interferon-β. Biochemical liver enzyme test results are either displayed as units per liter (U/l, i.e., ALP, ALT and AST) or milligrams per deciliter (mg/dl, i.e., TBIL). These plots display the interquartile range, with the medians depicted as black lines and whiskers indicating 1.5 times the interquartile range. Dashed lines indicate the upper limit of normal values for the respective biochemical liver test. Linear regression was performed using an additive genetic model for highest values, adjusted for age at biochemical liver test date, sex, and the first two principal components.



Supplementary Figure 4

Multi-tissue eQTL analyses for rs2205986 and *IRF6* from the GTEx Project.

m values represent the posterior probability for an eQTL effect (*PLoS Genet.* **8**, e1002555, 2012), and tissues with values >0.9 (i.e., cerebellum, tibial nerve, aorta, cerebellar hemisphere, whole blood and frontal cortex) are predicted to be eQTLs.

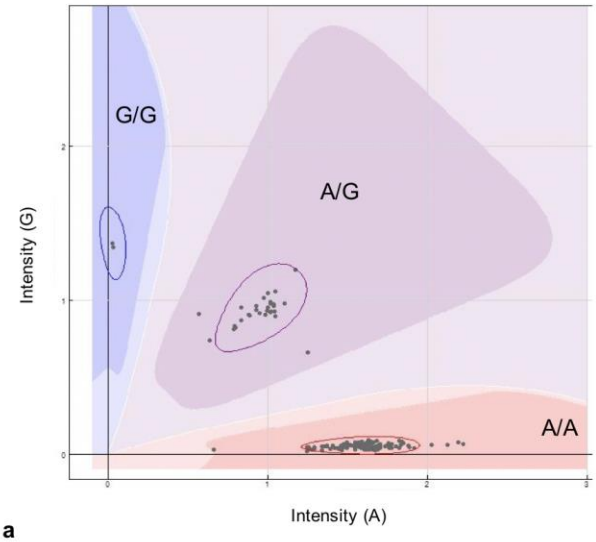


Model	Area Under the Curve	95% Confidence Interval	*P*-value vs. Clinical Model
Clinical	0.61	0.52-0.70	-
Clinical and Genomic	0.72	0.64-0.81	0.0039

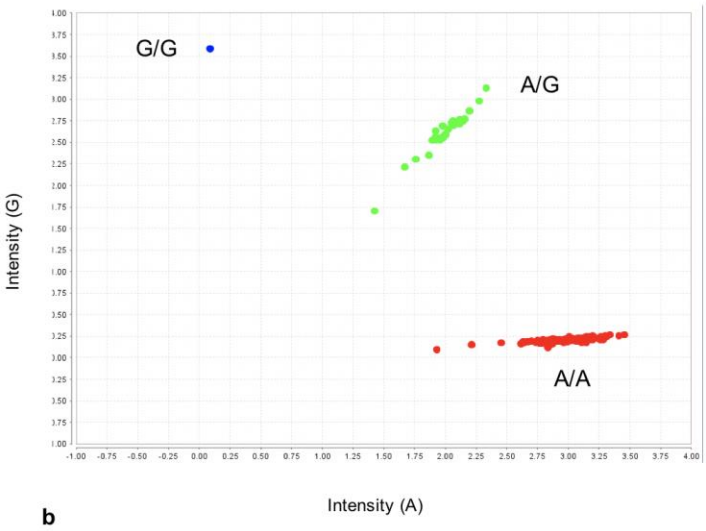
Supplementary Figure 5

Receiver operating characteristic (ROC) curve of the clinical and clinical/genomic model for predicting interferon-β-induced liver injury in the combined cohort.

The area under the curves and corresponding 95% confidence intervals are presented for the combined cohort (*n* = 182). Clinical variables were age at start of IFN-β, IFN-β product and sex. The genomic variable included was rs2205986. The ROC curves of these two prediction models were compared using DeLong's test.


G/G
A/G
A/A
Intensity (G)
Intensity (A)

a


G/G
A/G
A/A
Intensity (G)
Intensity (A)

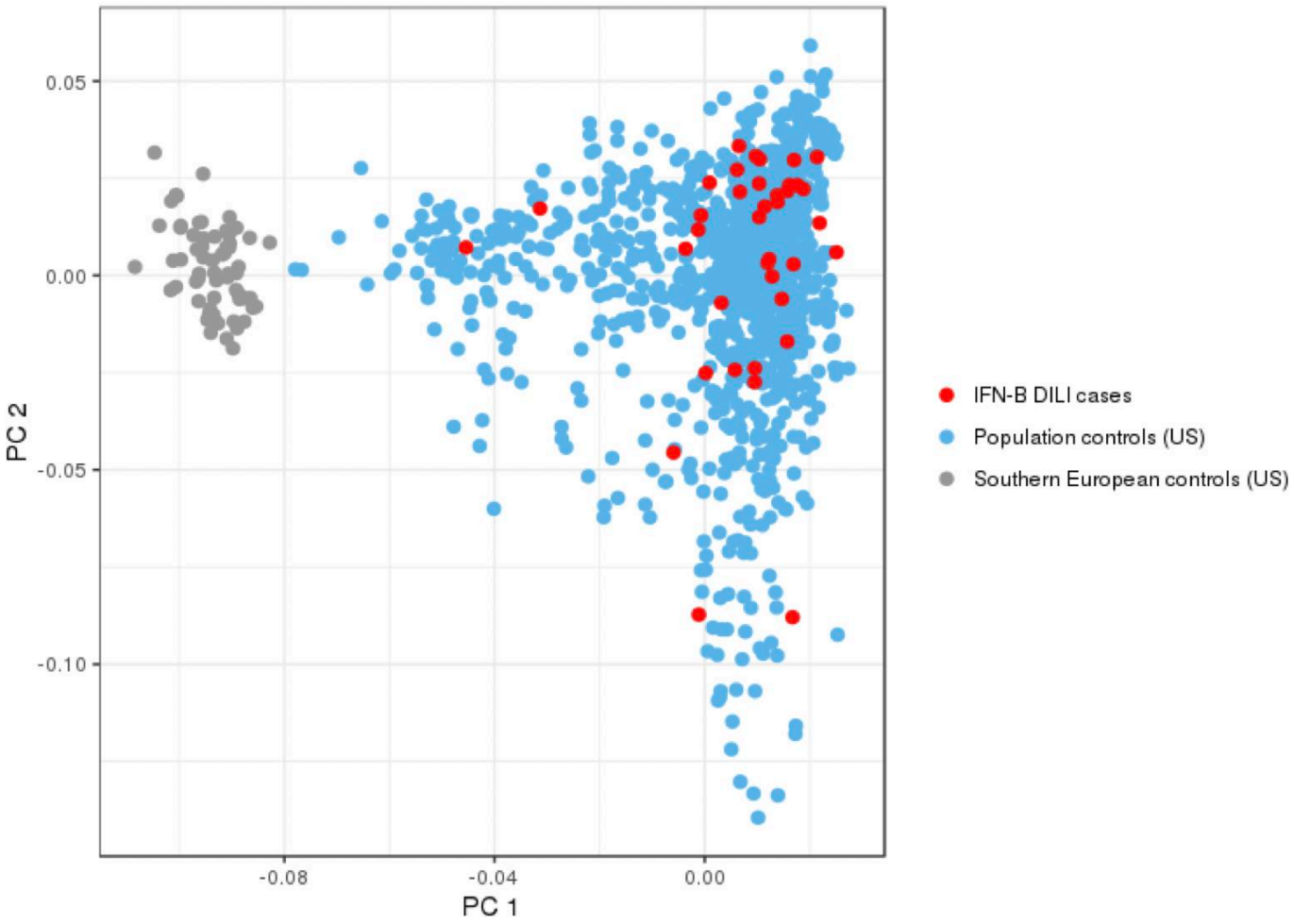
b

.

Supplementary Figure 6

Genotyping cluster plots of rs2205986, by genotyping method.

a, Genotyped by the Illumina MEGA array, plot generated using Illumina GenomeStudio software (*n* = 170 stage 1 and 10 stage 2 samples, all ancestries). One sample that presented as a G/G genotype was removed during quality control owing to low array call rate. **b**, TaqMan rs2205986 allelic discrimination plot generated using Thermo Fisher TaqMan Genotyper software (Methods) in stage 1 of European-ancestry patients (*n* = 151).


0.05
0.00
-0.05
-0.10
PC 2
-0.08
-0.04
0.00
PC 1
IFN-B DILI cases
Population controls (US)
Southern European controls (US)

Supplementary Figure 7

Principal-component analysis of disease-matched population controls.

Multiple sclerosis population controls from the United States ascertained to be of northern European genetic ancestry (blue points, *n* = 1,319) were included. Cases used in stage 1 analyses (red points) and excluded southern European samples (gray points) are included for reference.

Supplementary Table 1: **Assessment of clinical and demographic characteristics of multiple sclerosis patients exposed to IFN-β therapy in each stage of analysis**

Variable	Stage one (*n*=151)			Stage two (*n*=31)			Combined [Stage one + two] (*n*=182)		
	Cases (*n*=38)	Controls (*n*=113)	*P*[a]	Cases (*n*=18)	Controls (*n*=13)	*P*[a]	Cases (*n*=56)	Controls (*n*=126)	*P*[a]
Female, no. (%)	30 (78.9)	103 (91.2)	0.078[b]	15 (83.3)	12 (92.3)	0.462[c]	45 (80.4)	115 (91.3)	0.051[b]
Age[d], median y (IQR)	40.5 (30.8, 50.3)	40 (34, 47.5)	0.857[e]	38.5 (10.2)	29.9 (8.9) (*n*=12)	**0.026**[f]	40 (31.3, 47.8)	40 (33, 46.5) (*n*=125)	0.978[e]
Relapsing remitting course[d], no. (%)	34 (89.5)	111 (98.2)	**0.035**[c]	11 (100) (*n*=11)	12 (92.3)	0.347[c]	45 (91.8) (*n*=49)	123 (97.6)	0.080[c]
IFN-β product, no. (%)			0.195[c]			0.105[c]			0.464[c]
1a IM (30 mcg once weekly)	1 (2.6)	12 (10.6)		7 (38.9)	10 (76.9)		8 (14.3)	22 (17.5)	
1a SC (22 mcg 3x weekly)	5 (13.2)	13 (11.5)		0	0		5 (8.9)	13 (10.3)	
1a SC (44 mcg 3x weekly)	14 (36.8)	32 (28.3)		5 (27.8)	1 (7.7)		19 (33.9)	33 (26.2)	
1b (250 mcg every other day)	17 (44.7)	56 (49.6)		6 (33. 3)	2 (15.4)		23 (41.1)	58 (46.0)	
Pegylated 1a SC[g]	1 (2.6)	0		0	0		1 (1.8)	0	
BMI (kg/m^2), median (IQR)	26.6 (22.5, 30.2)	25.6 (22.5, 30.2)	0.973[e]	26.2 (21.3, 29) (*n*=7)	25.5 (25.5) (*n*=1)	0.408[e]	26.6 (22.0, 29.5) (*n*=45)	25.6 (22.5, 30.1), (*n*=114)	0.919[e]
Concomitant hepatotoxic medication use, no. (%)[h]	31 (81.6)	87 (77.0)	0.654[b]	16 (88.9)	10 (76.9)	0.370[c]	47 (83.9)	97 (77.0)	0.287[b]

[a]*P*-values represent cases compared to controls, bold indicates *P*<0.05; all *P*-values were 2-tailed. [b]Pearson's chi-square test, [c]Fisher Exact Test, [d]At the start of IFN-β therapy, [e]Mann-Whitney *U* Test, [f]Student's *t* Test; mean (±standard deviation) are reported for these variables, [9]Dose/frequency blinded as part of a clinical trial. [h]Hepatotoxic medication taken during the IFN-β exposure period, with hepatotoxicity defined by an international collaboration on Drug-Induced Liver Injury.[49] IQR: interquartile range, IM: intramuscular, SC: subcutaneous, BMI: body mass index.

Supplementary Table 2: **Genotype frequencies for rs2205986 in cases of liver injury due to IFN-β and controls that were IFN-β tolerant**

	Stage one (*n*=151)		Stage two (*n*=31)		Combined [Stage one + two] (*n*=182)	
Genotype	**Case (%)**	**Control (%)**	**Case (%)**	**Control (%)**	**Case (%)**	**Control (%)**
G/G	1 (2.6)	0	0	0	1 (1.8)	0
A/G	16 (42.1)	8 (7.1)	6 (33.3)	0	22 (39.3)	8 (6.4)
A/A	21 (55.3)	105 (92.9)	12 (66.7)	13 (100)	33 (58.9)	118 (93.6)
Total	38 (100)	113 (100)	18 (100)	13 (100)	56 (100)	126 (100)

Supplementary Table 3: **Genome-wide association results of the analyses between multiple sclerosis cases with IFN-β-induced liver injury and IFN-β tolerant multiple sclerosis controls for variants that passed the stage one significance threshold but did not replicate in stage two**

Genetic Variant Information				Pharmacogenomic Analyses			Logistic Regression (Additive)[a]		
Variant	**Nearest gene**	**Position[b]**	**Function**	**Study stage**	**MAF cases**	**MAF controls**	***P*-value**	**Odds ratio**	**95% CI**
rs7227310[c]	*CBLN2*	18: 70,098,428	Intergenic	Stage one	0.08	0.30	2.5×10^{-7}	0.2	0.1-0.5
				Stage two	0.17	0.25	0.42	0.6	0.2-2.1
				Stage one & two	0.11	0.29	2.2×10^{-5}	0.3	0.1-0.6
rs72663851[d]	*YTHDF2*	1: 29,059,972	Upstream gene	Stage one	0.07	0.29	6.7×10^{-7}	0.2	0.1-0.5
				Stage two	0.14	0.13	0.26	1.1	0.2-5.2
				Stage one & two	0.09	0.27	8.1×10^{-5}	0.3	0.1-0.6

[a]Logistic regression was performed in stage one (adjusted for MS disease course) and two (adjusted for age). [b]GRCh37 assembly position (chromosome:base pair). [c]Perfect linkage disequilibrium (r^2=1) with rs7242683. [d]Perfect linkage disequilibrium (r^2=1) with rs61787566, rs61787567 and rs17340752. CI: confidence interval, MAF: minor allele frequency.

Supplementary Table 4: Genetic association analysis of previously reported associations with *HLA*-alleles and drug-induced liver injury

Drug	*HLA*-allele	Original study cohorts (none specifically enrolled people with MS)					Current study (38 MS cases and 113 MS controls)				
		Sample size	*P*-value	OR	95% CI	Reference	*P*-value[c]	OR	95% CI	MAF (Cases)	MAF (Controls)
Ximelagatran	*DRB1*07*	74 cases 130 treated controls	7.3×10^{-8}	4.4	2.2-9.9	PMID: 17505501	0.84	1.1	0.5-2.5	0.13	0.12
Flucloxacillin	*DRB1*07:01*	51 cases 64 treated controls	1.6×10^{-6}	7.2	3.2-16.5	PMID: 19483685	0.84	1.1	0.5-2.5	0.13	0.12
Flucloxacillin	*B*57:01*	51 cases 282 population controls	8.7×10^{-33}	45.0	19.4-105.0	PMID: 19483685	NA[a]	NA[a]	NA[a]	0.00	0.02
Lumiracoxib	*DRB1*15:01*	41 cases 176 treated controls	2.8×10^{-10}	5.3	3.0–9.2	PMID: 20639878	0.68	1.2	0.5–2.7	0.14	0.13
Lapatinib	*DQA1*02:01*	37 cases 286 treated controls	0.03	2.6	1.1-5.7	PMID: 21245432	0.74	0.8	0.2-3.2	0.04	0.04
Amoxicillin-clavulanate	*DRB1*15:01*	201 cases 532 population controls	3.5×10^{-11}	2.8	2.1-3.8	PMID: 21570397	0.68	1.2	0.5–2.7	0.14	0.13
Antituberculosis treatment	*DQA1*01:02*	56 cases 290 treated controls	<0.01	4.0	1.1–14.3	PMID: 12359646	0.93	1.0	0.6-1.7	0.42	0.41
Antituberculosis treatment	*DQB1*05*	88 cases 88 treated controls	0.03	5.3	1.1–24.6	PMID: 25250564	0.48	1.3	0.6-2.8	0.12	0.10
Ticlopidine	*A*33:03*	22 cases 85 treated controls	1.2×10^{-5}	13.0	4.4-38.6	PMID: 17339877	0.09	NA[b]	NA[b]	0.01	0.00
Licensed drugs causing DILI without previously reported genetic risk factors	*A*33:01*	862 cases 10,588 population-matched controls	8.0×10^{-8}	2.6	1.8-3.7	PMID: 28043905	0.09	NA[b]	NA[b]	0.01	0.00
Minocycline	*B*35:02*	25 cases 6,835 population-matched controls	2.5×10^{-8}	29.6	7.8-89.8	PMID: 28323125	0.41	3.3	0.2-54.8	0.01	0.00

CI: Confidence interval, DILI: Drug-induced liver injury; OR: Odds ratio, MAF: Minor allele frequency; NA: Not applicable.
[a]Regression failed as none of the MS cases carried the variant. [b]Only one heterozygous individual (case) was observed, therefore an accurate OR could not be determined.
[c]*P* values were based on association results obtained using logistic regression analyses (additive model) including MS disease course as a covariate. $P<0.05$ was considered significant.

Supplementary Table 5: **Top variants within the chromosome 1q32.2 region from the pharmacogenomic association analyses for IFN-β-induced liver injury**

SNP	Position	Reference	Alternate	Annotation	MAF cases	MAF controls	OR	OR lower	OR upper	P^a	P^b (conditioned rs2205986)	LD (r^2) rs2205986
rs78270913	1:210068504	GTTGTTGTT	G	Intergenic	0.24	0.07	4.2	2.0	8.6	3.4E-05	0.52	0.54
rs72741075	1:210074833	C	T	Intergenic	0.23	0.06	4.4	2.1	9.3	3.1E-05	0.12	0.78
rs12086774	1:210079167	G	C	Intergenic	0.24	0.07	4.4	2.1	9.3	3.0E-05	0.12	0.78
rs2357211	1:210106611	A	G	Upstream gene	0.26	0.08	3.9	1.9	7.8	2.4E-05	0.95	0.62
rs2205986	1:210116112	G	A	Intron	0.24	0.04	8.5	3.5	20.4	1.9E-07	1	1.00
rs1016479	1:210119135	A	C	Intron	0.26	0.08	3.9	1.9	7.8	2.4E-05	0.95	0.62
rs7518683	1:210122787	A	G	Intron	0.26	0.08	3.9	2.0	7.9	2.0E-05	0.93	0.64
rs719034	1:210135806	A	G	Intron	0.26	0.08	3.9	1.9	7.8	2.4E-05	0.95	0.62
rs2205990	1:210145771	C	A	Intron	0.29	0.09	4.0	2.0	7.8	1.2E-05	0.48	0.55
rs149582815	1:210148017	A	T	Intron	0.26	0.08	3.9	1.9	7.8	2.4E-05	0.95	0.62
rs61820393	1:210161119	A	G	Intron	0.29	0.09	4.0	2.0	7.8	1.2E-05	0.48	0.55
rs227214	1:210179775	T	C	Intron	0.29	0.09	4.0	2.0	7.8	1.3E-05	0.48	0.55
rs227215	1:210180530	G	T	Intron	0.29	0.09	4.0	2.0	7.8	1.3E-05	0.48	0.55
rs138320559	1:210198206	TAGA	T	Intron	0.29	0.09	4.0	2.0	7.7	1.3E-05	0.48	0.55
rs1099864	1:210202253	A	G	Intron	0.29	0.10	3.8	1.9	7.3	2.1E-05	0.55	0.53
rs553593586	1:210203473	A	AGTTTTTTTT	Intron	0.29	0.10	3.8	1.9	7.3	2.1E-05	0.55	0.53
rs227192	1:210210240	T	C	Intron	0.29	0.10	3.8	1.9	7.3	2.1E-05	0.55	0.53
rs2743892	1:210227209	G	A	Intron	0.28	0.09	3.7	1.9	7.3	4.2E-05	0.65	0.51
rs227208	1:210241176	C	G	Intron	0.29	0.09	4.0	2.0	7.7	1.3E-05	0.49	0.55
rs200837868	1:210248922	CCTGCCT	C	Intron	0.29	0.09	4.0	2.0	7.7	1.3E-05	0.49	0.55
rs227185	1:210252888	G	A	Intron	0.29	0.09	4.0	2.0	7.7	1.3E-05	0.49	0.55
rs846547	1:210260460	G	A	Intron	0.29	0.09	3.9	2.0	7.7	1.3E-05	0.49	0.55
rs2307890	1:210267893	TGAA	T	Inframe deletion	0.28	0.09	3.9	2.0	7.7	2.8E-05	0.41	0.48
rs2494187	1:210273886	A	T	Intron	0.25	0.08	4.1	2.0	8.5	2.4E-05	0.10	0.31
rs2494188	1:210287723	C	T	Intron	0.25	0.08	4.1	2.0	8.5	2.4E-05	0.10	0.31

rs181615795	1:210295383	T	C	Intron	0.25	0.08	3.9	1.9	7.9	4.4E-05	0.19	0.35
rs2451693	1:210296914	A	C	Intron	0.25	0.08	4.1	2.0	8.4	2.5E-05	0.10	0.31
rs2484029	1:210298944	C	A	Intron	0.25	0.08	4.1	2.0	8.4	2.5E-05	0.10	0.30
rs923561	1:210302595	G	A	Intron	0.25	0.08	4.1	2.0	8.4	2.5E-05	0.10	0.30
rs17188183	1:210334081	T	C	Synonymous	0.24	0.08	3.9	1.9	8.1	4.9E-05	0.14	0.31

Logistic regression was adjusted for either [a]MS disease course or [b]MS disease course and rs2205986.
CI: Confidence interval, DILI: Drug-induced liver injury; LD, linkage disequilibrium, OR: Odds ratio, MAF: Minor allele frequency